Department of Mathematics
and Statistics
The University of South Carolina
Columbia, South Carolina 29208

AFOSR-TR- 84-1150

5

A KERNEL TYPE ESTIMATOR OF A QUANTILE FUNCTION FROM RIGHT-CENSORED DATA *

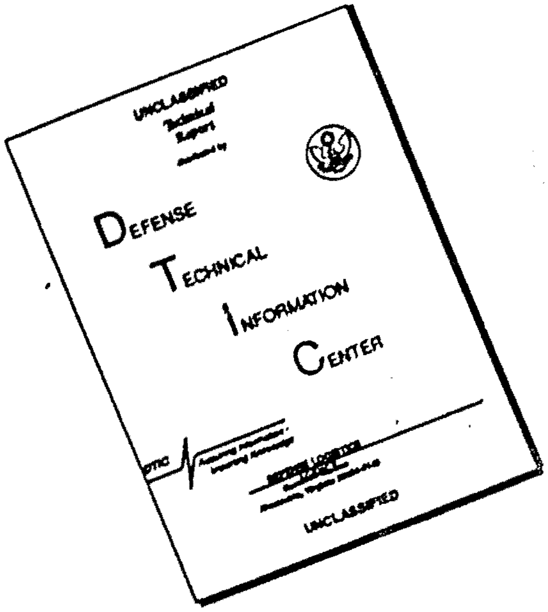
by

W. J. Padgett

University of South Carolina

Statistics Technical Report No. 101
62G05-10

DISCLAIMER NOTICE



THIS DOCUMENT IS BEST QUALITY AVAILABLE. THE COPY FURNISHED TO DTIC CONTAINED A SIGNIFICANT NUMBER OF PAGES WHICH DO NOT REPRODUCE LEGIBLY.

A KERNEL TYPE ESTIMATOR OF A QUANTILE FUNCTION FROM RIGHT-CENSORED DATA *

by

W. J. Padgett

University of South Carolina
Statistics Technical Report No. 101
62G05-10

November 1984

Department of Mathematics and Statistics
University of South Carolina
Columbia, South Carolina 29208

* Research supported by the United States Air Force Office of Scientific Research and Army Research Office under Grant No. AFOSR-84-0156.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) Statistics Technical Report No. 101 (62G05-10)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 84-1150
6a. NAME OF PERFORMING ORGANIZATION University of South Carolina 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Air Force Office of Scientific Research
6c. ADDRESS (City, State and ZIP Code) Department of Mathematics & Statistics Columbia SC 29208	7b. ADDRESS (City, State and ZIP Code) Directorate of Mathematical & Information Sciences, Bolling AFB DC 20332-6448
8a. NAME OF FUNDING/SPONSORING ORGANIZATION AFOSR 8b. OFFICE SYMBOL (If applicable) NM	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-84-0156
8c. ADDRESS (City, State and ZIP Code) Bolling AFB DC 20332-6448	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO.: 61102F PROJECT NO.: 2304 TASK NO.: A5 WORK UNIT NO.:

11. TITLE (Include Security Classification)
A KERNEL TYPE ESTIMATOR OF A QUANTILE FUNCTION FROM RIGHT-CENSORED DATA

12. PERSONAL AUTHOR(S)
W.J. Padgett

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Yr., Mo., Day)	15. PAGE COUNT
Technical	FROM ______ TO ______	1 NOV 84	22

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	Random censorship; product-limit quantile function; Kernel estimation; median survival time; nonparametric quantile function.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Based on right-censored data from a lifetime distribution F_o, a kernel type estimator of the quantile function $Q^o(p)=\inf\{t: F_o(t)\geq p\}$, $0\leq p\leq 1$, is proposed. The estimator is defined by $Q_n(p) = h_n^{-1}\int_0^1 \hat{Q}_n(t)K((t-p)/h_n)dt$, which is smoother than the usual product-limit quantile function $\hat{Q}_n(p)=\inf\{t: \hat{F}_n(t)\geq p\}$, where $\hat{F}_n$ denotes the product-limit estimator of F_o from the censored sample. Under the random censorship model and general conditions on h_n, K, and F_o, it is shown that $Q_n(p)$ is strongly consistent. A simple approximation to Q_n is also shown to be asymptotically equivalent to Q_n with probability one. A small Monte Carlo simulation study shows that for several values of the bandwidth h_n, Q_n performs better than $\hat{Q}_n$ in the sense of estimated mean squared errors. The estimator is illustrated by an application to data from a mechanical-switch life test.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL MAJ Brian W. Woodruff	22b. TELEPHONE NUMBER (Include Area Code) (202) 767- 5027 22c. OFFICE SYMBOL NM

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

84 12 13 020

ABSTRACT

Based on right-censored data from a lifetime distribution F_0, a kernel type estimator of the quantile function $Q^0(p) = \inf\{t: F_0(t) \geq p\}$, $0 \leq p \leq 1$, is proposed. The estimator is defined by $Q_n(p) = h_n^{-1} \int_0^1 \hat{Q}_n(t)K((t-p)/h_n)dt$, which is smoother than the usual product-limit quantile function $\hat{Q}_n(p) = \inf\{t: \hat{F}_n(t) \geq p\}$, where $\hat{F}_n$ denotes the product-limit estimator of F_0 from the censored sample. Under the random censorship model and general conditions on h_n, K, and F_0, it is shown that $Q_n(p)$ is strongly consistent. In addition, an approximation to Q_n is shown to be asymptotically equivalent to Q_n with probability one. A small Monte Carlo simulation study shows that for several values of the bandwidth h_n, Q_n performs better than $\hat{Q}_n$ in the sense of estimated mean squared errors. The estimator is illustrated by an application to data from a mechanical-switch life test.

Key Words: Random censorship; Product-limit quantile function; Kernel estimation; Median survival time estimation; Nonparametric quantile estimation.

1. INTRODUCTION

Arbitrarily right-censored data arise naturally in industrial life testing and medical follow-up studies. In these situations it is important to be able to obtain nonparametric estimates of various characteristics of the survival function S. Based on such right-censored data, Kaplan and Meier ~~(1958)~~ gave the nonparametric maximum likelihood estimator of S, called the product-limit estimator, and, among others, Reid (1981) has proposed methods of estimating the median survival time from the product-limit estimator. Recently, Nair (1984) studied the problem of confidence bands for the survival function obtained from the product-limit estimator. Also, Padgett and McNichols ~~(1984)~~ and McNichols and Padgett (1984) have discussed estimation of a density for the survival distribution based on right-censored data.

One characteristic of the survival distribution that is of interest is the quantile function, which is useful in reliability and medical studies. For any probability distribution function G, the quantile function is defined by $Q(p) \equiv G^{-1}(p) \equiv \xi_p = \inf\{x: G(x) \geq p\}$, $0 \leq p \leq 1$. In particular, $\xi_{0.5}$ is a median of G. For a random (uncensored) sample $Y_1,\ldots,Y_n$ from G, the sample quantile function $G_n^{-1}(p) = \inf\{x: G_n(x) \geq p\}$, $0 \leq p \leq 1$, has been used to estimate ξ_p, where $G_n(x)$ denotes the sample distribution function. Note that $G_n^{-1}(p) = Y_{([np])}$, the [np]th order statistic among $Y_1,\ldots,Y_n$, where $[\cdot]$ denotes the greatest integer function. Csörgő (1983) gives many of the known results concerning $G_n^{-1}(p)$. Also, Falk (1984) has recently studied the relative deficiency of the sample quantile with respect to kernel type estimators.

Other nonparametric estimators of the quantile function from uncensored data have been proposed which are smoother than the sample quantile function.

For example, Kaigh and Lachenbruch (1982) considered a "generalized sample quantile" obtained by averaging an appropriate subsample quantile over all subsamples of a given size. Also recently, Yang (1984) has studied the properties of kernel-type estimators of ξ_p which smooth the sample quantile function. Parzen (1979) had mentioned kernel estimators as a possible class of quantile estimators, but did not investigate their properties.

For arbitrarily right-censored data, Sander (1975) proposed estimation of ξ_p by the quantile function of the product-limit estimator of 1-S, and she and Cheng (1981) obtained some asymptotic properties of that estimator. Csörgő (1983) presented strong approximation results for this estimator.

The quantile function of the product-limit estimator is a step function with jumps corresponding to the uncensored observations. The purpose of this paper is to present a smoothed nonparametric estimator of the quantile function from arbitrarily right-censored data based on the kernel method. It will be shown that under general conditions this estimator, mentioned briefly by Parzen (1979, p. 119), is strongly consistent, and based on the results of a small Monte Carlo simulation study, performs better than the quantile function of the product-limit estimator in the sense of smaller mean squared error. In particular, better estimates of the median survival time are obtainable. In addition, an approximation to the kernel estimator will be shown to be almost surely asymptotically equivalent to it under certain conditions. Finally, estimates of the quantile function from the randomly right-censored data given by Nair (1984) are presented as an illustration.

2. ARBITRARILY RIGHT-CENSORED DATA

Let $X_1^o,\ldots,X_n^o$ denote the true survival times of n items or individuals which are censored on the right by a sequence $U_1,U_2,\ldots,U_n$, which in general may be either constants or random variables. It is assumed that the X_i^o's are nonnegative independent identically distributed random variables with common unknown distribution function F_0 and unknown quantile function $Q^o(p) \equiv \xi_p^o \equiv \inf\{t: F_0(t) \geq p\}$, $0 \leq p \leq 1$. Also, $Q^o(p)$ is sometimes denoted by $F_0^{-1}(p)$.

The observed right-censored data are denoted by the pairs (X_i,Δ_i), $i=1,\ldots,n$, where

$$X_i = \min\{X_i^o,U_i\}, \qquad \Delta_i = \begin{cases} 1 & \text{if } X_i^o \leq U_i \\ 0 & \text{if } X_i^o > U_i \,. \end{cases}$$

Thus, it is known which observations are times of failure or death and which ones are censored or loss times. The nature of the censoring depends on the U_i's. (i) If $U_1,\ldots,U_n$ are fixed constants, the observations are time-truncated. If all U_i's are equal to the same constant, then the case of Type I censoring results. (ii) If all $U_i = X_{(r)}^o$, the rth order statistic of $X_1^o,\ldots,X_n^o$, then the situation is that of Type II censoring. (iii) If $U_1,\ldots,U_n$ constitute a random sample from a distribution H (usually unknown) and are independent of $X_1^o,\ldots,X_n^o$, then (X_i,Δ_i), $i=1,2,\ldots,n$, is called a randomly right-censored sample.

The random censorship model (iii) is assumed for the results presented here. For this model, $\Delta_1,\ldots,\Delta_n$ are independent Bernoulli random variables, and the distribution function F of each X_i, $i=1,\ldots,n$, is given by $F = 1 - (1-F_0)(1-H)$.

Based on the censored sample (X_i,Δ_i), $i=1,2,\ldots,n$, a popular estimator of the survival function $1-F_0(t)$ at $t \geq 0$ is the product-limit estimator, proposed by Kaplan and Meier (1958) as the "nonparametric maximum likelihood estimator". Efron (1967) showed that this estimator, defined next, is "self-consistent". Let (Z_i,Δ_i'), $i=1,\ldots,n$, denote the ordered X_i's along with their corresponding Δ_i's. A value of the censored sample will be denoted by the corresponding *lower* case letters (x_i,δ_i) and (z_i,δ_i') for the unordered and ordered sample, respectively. Then the product-limit estimator of $1-F_0(t)$ is defined by

$$\hat{P}_n(t) = \begin{cases} 1, & 0 \leq t \leq Z_1, \\ \prod_{i=1}^{k-1} \left(\frac{n-i}{n-i+1}\right)^{\delta_i'}, & Z_{k-1} < t \leq Z_k,\ k=2,\ldots,n \\ 0, & Z_n < t. \end{cases}$$

Denote the product-limit estimator of $F_0(t)$ by $\hat{F}_n(t) = 1 - \hat{P}_n(t)$, and let s_j denote the jump of $\hat{P}_n$ (or $\hat{F}_n$) at Z_j, that is

$$s_j = \begin{cases} 1 - \hat{P}_n(Z_2), & j=1 \\ \hat{P}_n(Z_j) - \hat{P}_n(Z_{j+1}), & j=2,\ldots,n-1 \\ \hat{P}_n(Z_n), & j=n. \end{cases}$$

Note that $s_j = 0$ if and only if $\delta_j' = 0$, $j < n$, that is, if Z_j is a censored observation. Also, denote $S_i \equiv \hat{F}_n(Z_i) = \sum_{j=1}^{i} s_j$, $i=1,2,\ldots,n$.

The product-limit estimator has played a central role in the analysis of censored survival data (Miller, 1981). Its properties have been studied by many authors, for example, Breslow and Crowley (1974), Földes and Rejtő (1981),

Földes, Rejtö and Winter (1980), and Gill (1983).

Based on randomly right-censored data, it is natural to estimate the quantile function $Q^o(p)$ by the product-limit (PL) quantile function $\hat{Q}_n(p) \equiv \hat{\xi}_p \equiv \inf\{t: \hat{F}_n(t) \geq p\}$. Cheng (1981) obtained asymptotic normality results for $\hat{\xi}_p$ and gave an asymptotic expression for $\hat{\xi}_p$ in terms of ξ_p^o, $\hat{F}_n$, and f_0, a density function of F_0. Csörgő (1983) presented strong approximation theorems for the PL quantile process $\hat{Q}_n$.

3. THE QUANTILE ESTIMATOR

In this section the kernel estimator of $Q^o(p)$, $0 \leq p \leq 1$, from the randomly right-censored observations (X_i,Δ_i), $i=1,\ldots,n$, will be defined. Similar to Yang's (1984) estimators for the uncensored case, an approximation which is often easier to compute will be given. First, some assumptions and notation concerning the kernel, the bandwidth sequence, and the lifetime and censoring distributions will be listed.

Let $\{h_n\}$ be a "bandwidth" sequence of positive numbers such that

(h.1) $\qquad h_n \to 0$ as $n \to \infty$.

Let K be a real-valued function defined on $(-\infty,\infty)$ such that

(K.1) $\qquad K(x) \geq 0$, all real numbers x,

(K.2) $\qquad \int_{-\infty}^{\infty} K(x)dx = 1$,

(K.3) $\qquad K$ has finite support, that is, $K(x) = 0$ for $|x| > c$ for some $c > 0$,

(K.4) $\qquad \sup_x |K(x)| < \infty$, that is, K is bounded,

(K.5) K is symmetric about zero, and

(K.6) K satisfies a Lipschitz condition, that is, there exists a constant Γ such that for all x,y,

$$|K(x) - K(y)| \le \Gamma |x - y|.$$

Notice that conditions (K.1) - (K.2) simply say that K must be a probability density function. Also, assume that the lifetime distribution F_0 and the censoring distribution H are such that

(F.1) F_0 is continuous with density function f_0,

(F.2) f_0 is continuous at $\xi_p^o = Q^o(p)$ and $f_0(\xi_p^o) > 0$,

(F.3) F_0 has a finite mean, and

(F.4) $H(T_{F_0}) < 1$, where $T_{F_0} = \sup\{t: F_0(t) < 1\}$.

It should be noted that the conditions (F.1) - (F.4) are not prohibitive and are similar to those assumed by Cheng (1981). Condition (F.4) is usually required for asymptotic results with random right-censorship and guarantees that observations can be obtained from the entire support of the distribution F_0. The conditions (h.1), (K.3 - K.6), and (F.2) - (F.4) are required for the asymptotic results of Section 4.

Now, for $0 \le p \le 1$, define the kernel type quantile function estimator

$$Q_n(p) = h_n^{-1} \int_0^1 \hat{Q}_n(t) K((t-p)/h_n)dt$$

$$= h_n^{-1} \sum_{i=1}^{n} Z_i \int_{S_{i-1}}^{S_i} K((t-p)/h_n)dt, \qquad (3.1)$$

where $S_i = \hat{F}_n(Z_i)$. It should be noted that only those Z_i which are uncensored appear in the sum (3.1) since

$$\int_{S_{i-1}}^{S_i} K((t-p)/h_n)dt = \begin{cases} 0, & \text{if } Z_i \text{ is censored} \\ h_n[K^*((S_i - p)/h_n) - K^*((S_{i-1} - p)/h_n], & \\ \quad \text{if } Z_i \text{ is uncensored,} & \end{cases}$$

where K^* denotes the cumulative distribution function of K.

An approximation to the estimator (3.1) can be obtained by noticing that the derivative of K^* at $(S_i - p)/h_n$ is approximated by

$$(h_n/s_i)\ [K^*((S_i - p)/h_n) - K^*((S_{i-1} - p)/h_n] \approx K((S_i - p)/h_n).$$

Hence, when $S_i - S_{i-1}$ is small, (3.1) is approximately equal to

$$Q_n^*(p) = h_n^{-1} \sum_{i=1}^{n} Z_i s_i\ K((S_i - p)/h_n). \tag{3.2}$$

Again, since $s_i = 0$ when $\delta_i = 0$, $i < n$, only the uncensored observations explicitly appear in the sum (3.2).

In the case of no censoring, (3.1) and (3.2) reduce to the kernel estimators of Yang (1984). He has shown that his estimators are asymptotically equivalent in mean square and obtained rates of convergence for the variance and bias. Due to the censoring, similar results for the variance, bias, and mean square consistency of (3.1) and (3.2) seem to be difficult, if not impossible, to obtain under general conditions on F_0 and H. Some asymptotic results, however, have been obtained under reasonable conditions and are presented in the next section.

4. ASYMPTOTIC RESULTS

Here two asymptotic results for Q_n and Q_n^* will be obtained. First, the almost sure consistency of the kernel estimator $Q_n(p)$ is stated. The proofs of Theorems 1 and 2 are given in the Appendix.

Theorem 1. Suppose the conditions (h.1), (K.1) - (K.6), and (F.1) - (F.4) hold. If $(\log \log n/n)^{3/4} h_n^{-1} \to 0$ as $n \to \infty$, then for each $0 \leq p \leq 1$, $Q_n(p) \to Q^o(p)$ as $n \to \infty$ with probability one.

The two estimators Q_n and Q_n^* can be shown to be asymptotically almost surely (uniformly in p) equivalent under general conditions. First, define $\hat{\mu}(t) \equiv \int_0^t \hat{P}_n(x)dx$, $t \geq 0$.

Theorem 2. Suppose F_0 and H are continuous and that (h.1), (K.1), (K.2), (K.6), (F.3), and (F.4) hold. If $\lim_n \sup \hat{\mu}(Z_n) < \infty$ with probability one, then

$$P[|Q_n^*(p) - Q_n(p)| = O(h_n^{-2}(\log \log n/n)^{\frac{1}{2}})] = 1.$$

Thus, if $h_n^{-2}(\log \log n/n)^{\frac{1}{2}} \to 0$ as $n \to \infty$, then under the above conditions Q_n^* and Q_n are (uniformly in p) asymptotically equivalent with probability one.

It should be remarked that $\lim_n \sup \hat{\mu}(Z_n) < \infty$ almost surely under the conditions given by Susarla and Van Ryzin (1980), for example.

Also, if $h_n \simeq D\, n^{-b}$ for $0 < b < \frac{1}{2}$ and some positive constant D, the condition that $h_n^{-2}(\log\log n/n)^{\frac{1}{2}} \to 0$ holds. It seems to be quite difficult to obtain the exact mean squared error of Q_n or Q_n^* and to be able to choose $\{h_n\}$ to minimize this mean squared error or to choose an optimal $\{h_n\}$ in some other sense. Some simulation results presented in the next section indicate a range of possible h_n values for which the mean squared errors of Q_n (and Q_n^*) are less than those of the PL quantile estimator.

5. SOME SIMULATION RESULTS AND AN EXAMPLE

A small Monte Carlo study was performed in order to provide some small-sample comparisons of Q_n and Q_n^* with the PL quantile estimator, and with each other, in the sense of mean squared errors. The study also provides some insight into the choice of reasonable values for h_n which might be used in practice. The random censorship model with $F_0(t) = 1 - \exp(-t)$ and $H(t) = 1 - \exp(-\lambda t)$ was used with λ chosen to give 50% censoring or 30% censoring as in Reid (1981). The ratios of the mean squared error of $\hat{Q}_n(p)$ to the mean squared errors of the smoothed estimators $Q_n(p)$ and $Q_n^*(p)$ were computed for various $0 < p < 1$ and sample sizes $n = 50$ and 100. For each case, 1000 censored samples were generated using the uniform random number generator GGUBS in the International Mathematical and Statistical Libraries (1982) on a DEC VAX 11-750 computer. The standard errors of the simulated mean squared errors ranged from 10^{-1} to 10^{-4}.

Table 1 shows some of the results for the triangular kernel $K(x) = 1 - |x|$, $|x| \leq 1$, which satisfies the conditions (K.1) - (K.6) of Section 3. The simulations were run for values of $h_n = 0.01\ (0.02)\ 0.61$. For the estimator $Q_n(p)$, for each value of p listed there is an h_n which gives smaller estimated mean squared error than the PL quantile estimator. In particular, this is true for several h_n values for the median estimators $Q_n(0.5)$ and $\hat{Q}_n(0.5)$. The approximation $Q_n^*(p)$ performs well for several h_n values when $p \leq 0.5$, but not so for larger p. As would be expected for more severe censoring, the performance of either estimator at large values of p is not as good as for values near 0.5. Notice that h_n values of 0.09 to 0.13 appear to be best for $Q_n(p)$ over <u>all</u> p in Table 1(a) with n = 100, whereas for $Q_n^*(p)$ the h_n should be somewhat larger (0.15 to 0.21) for a good estimator over all p. Generally, the best h_n for $Q_n^*(p)$ is larger than that for $Q_n(p)$, indicating that Q_n^* requires slightly more smoothing than Q_n.

The results of Table 2 are for the uniform kernel $K(x) = 1$, $|x| \leq \frac{1}{2}$. This kernel does not satisfy condition (K.6), but the simulation results are quite similar except perhaps for the best choices of values of h_n.

It was mentioned in Section 4 that it is difficult if not impossible, to calculate in general the <u>exact</u> mean squared error of Q_n or Q_n^* for small n due to the right-censorship. Also, the mean square convergence (with a rate)

has not yet been obtained. Hence, to find an optimal h_n in the sense of minimum mean squared error of Q_n or Q_n^* seems to be quite difficult. These simulations, however, indicate reasonable ranges of h_n values which give small mean squared errors under the assumed models and censoring percentages.

As an example of the quantile estimators, the life test data for $n = 40$ mechanical switches reported by Nair (1984) are used. Two failure modes, A and B, were recorded and Nair (1984) estimated the survival function of mode A, assuming the random right-censorship model. Table 3 shows the 40 observations with the corresponding δ_i values ($\delta_i = 1$ indicates failure mode A and $\delta_i = 0$ denotes a censored value). There are seventeen uncensored observations, slightly more than 50% censoring. From Table 1(a) with $n = 50$, the values of h_n chosen for this example were 0.03 for $Q_n(p)$ and 0.15 for $Q_n^*(p)$. (Also, respective values of 0.05 and 0.19 were tried yielding similar estimates.) Figure 1 shows the estimates $Q_n(p)$ and $Q_n^*(p)$, calculated using the triangular kernel, along with the PL quantile function $\hat{Q}_n(p)$. Due to the large number of censored observations, the estimates for large p reflect the small estimated mean squared error ratios in Tables 1 and 2 for $p \geq 0.90$. In particular, the estimate Q_n^* is not very smooth for moderate to large p and could be smoothed more by taking larger h_n, say $h_n = 0.35$. However, as indicated by Table 1, the performance deteriorates for larger p with such h_n, and the estimate Q_n^* falls much below Q_n and $\hat{Q}_n$ for this data. The estimates of median lifetime are $Q_n(0.50) = 2.5478$, $Q_n^*(0.5) = 2.4354$, and $\hat{Q}_n(0.5) = 2.5480$.

TABLE 1. Ratios of MSEs with Triangular Kernel

(a) 50% Censoring

n = 100

p \ h_n		.03	.05	.07	.09	.11	.13	.15	.19	.21	.25	.31	.35	.41
.10	a.	1.16	1.18	1.31	1.34	1.43	1.40	1.51	1.40	1.27	1.17	0.84	0.65	0.38
	b.	1.29	1.37	1.52	1.55	1.66	1.68	1.85	1.75	1.66	1.58	1.15	0.88	0.50
.25	a.	1.04	1.07	1.09	1.12	1.16	1.21	1.16	1.23	1.22	1.17	1.15	1.08	0.93
	b.	1.17	1.20	1.22	1.27	1.31	1.35	1.35	1.40	1.44	1.43	1.50	1.47	1.40
.50	a.	1.04	1.07	1.10	1.12	1.14	1.17	1.14	1.16	1.15	1.10	1.03	0.89	0.73
	b.	1.14	1.24	1.34	1.34	1.39	1.47	1.46	1.50	1.61	1.72	1.74	1.85	1.83
.75	a.	1.07	1.19	1.23	1.41	1.41	1.31	1.20	1.43	1.34	1.31	1.60	1.85	2.64
	b.	0.33	0.63	1.16	1.53	1.51	1.69	1.27	1.47	1.31	1.30	2.40	2.66	2.65
.90	a.	1.11	1.13	1.23	1.30	1.40	1.51	1.45	1.14	0.94	0.92	0.50	0.51	0.37
	b.	0.10	0.16	0.21	0.22	0.37	0.64	0.79	0.76	0.70	0.76	0.45	0.48	0.36
.95	a.	1.02	1.02	0.88	0.69	0.57	0.49	0.41	0.34	0.29	0.28	0.22	0.22	0.19
	b.	0.09	0.10	0.37	0.55	0.57	0.54	0.49	0.40	0.33	0.31	0.24	0.24	0.20

n = 50

.10	a.	1.08	1.12	1.22	1.26	1.29	1.38	1.41	1.47	1.34	1.24	0.91	0.87	0.59
	b.	1.48	1.52	1.58	1.75	1.79	1.92	2.06	2.17	2.04	1.97	1.54	1.43	0.98
.25	a.	1.04	1.09	1.11	1.13	1.16	1.16	1.19	1.20	1.22	1.21	1.21	1.17	0.97
	b.	1.27	1.32	1.32	1.48	1.45	1.53	1.57	1.57	1.71	1.71	1.90	1.88	1.80
.50	a.	1.09	1.07	1.07	1.10	1.15	1.17	1.26	1.21	1.10	1.21	1.04	0.96	0.97
	b.	0.34	1.08	1.34	1.57	1.73	1.92	2.09	1.89	2.00	2.40	2.14	2.11	2.25
.75	a.	1.13	1.10	1.19	1.19	1.33	1.29	1.41	1.56	1.42	1.62	1.92	2.35	2.65
	b.	0.17	0.36	0.61	0.76	1.07	0.93	0.95	1.29	0.98	1.22	2.16	2.74	2.20
.90	a.	1.06	1.11	1.13	1.15	1.16	1.16	1.08	0.90	0.77	0.69	0.51	0.48	0.34
	b.	0.11	0.13	0.17	0.16	0.25	0.45	0.60	0.70	0.63	0.63	0.50	0.48	0.34
.95	a.	1.00	0.99	0.88	0.75	0.63	0.55	0.51	0.41	0.39	0.36	0.31	0.31	0.29
	b.	0.12	0.13	0.47	0.66	0.67	0.68	0.68	0.55	0.53	0.46	0.38	0.36	0.32

a = (MSE $\hat{F}_n^{-1}$)/(MSE Q_n), b = (MSE $\hat{F}_n^{-1}$)/(MSE Q_n^*)

TABLE 1. Ratios of MSEs with Triangular Kernel

(b) 30% Censoring

n = 100

p \ h_n		.03	.05	.07	.09	.11	.13	.15	.19	.21	.25	.31	.35	.41
.10	a.	1.10	1.15	1.24	1.26	1.32	1.34	1.41	1.37	1.19	1.11	0.80	0.62	0.35
	b.	1.22	1.30	1.39	1.44	1.50	1.55	1.66	1.64	1.49	1.42	1.02	0.78	0.43
.25	a.	1.04	1.08	1.09	1.13	1.14	1.17	1.17	1.20	1.21	1.22	1.22	1.09	0.94
	b.	1.09	1.16	1.17	1.22	1.22	1.25	1.28	1.29	1.35	1.38	1.43	1.31	1.19
.50	a.	1.04	1.05	1.07	1.09	1.10	1.11	1.12	1.16	1.09	1.11	1.06	0.94	0.74
	b.	1.10	1.11	1.15	1.19	1.20	1.21	1.22	1.26	1.24	1.29	1.29	1.22	1.05
.75	a.	1.02	1.15	1.10	1.10	1.08	1.11	1.12	1.05	1.00	0.79	0.82	0.98	1.41
	b.	1.02	1.38	1.27	1.33	1.35	1.40	1.43	1.53	1.57	1.50	1.39	1.46	1.68
.90	a.	1.11	1.23	1.34	1.21	1.23	1.67	1.71	1.73	1.21	1.05	0.55	0.52	0.36
	b.	0.69	0.97	1.25	1.15	1.34	2.07	1.72	1.61	1.10	0.98	0.53	0.50	0.35
.95	a.	1.23	1.27	1.61	1.66	1.18	0.94	0.69	0.51	0.40	0.34	0.27	0.26	0.21
	b.	0.32	0.38	1.18	1.50	1.08	0.94	0.70	0.53	0.42	0.36	0.28	0.27	0.22

n = 50

p \ h_n		.03	.05	.07	.09	.11	.13	.15	.19	.21	.25	.31	.35	.41
.10	a.	0.95	1.00	1.06	1.12	1.15	1.21	1.20	1.34	1.17	1.17	0.83	0.77	0.55
	b.	1.17	1.26	1.29	1.44	1.46	1.54	1.59	1.80	1.61	1.66	1.23	1.10	0.77
.25	a.	1.07	1.07	1.08	1.12	1.14	1.10	1.17	1.17	1.21	1.18	1.23	1.23	1.12
	b.	1.21	1.21	1.18	1.28	1.29	1.27	1.35	1.37	1.45	1.42	1.59	1.60	1.55
.50	a.	1.05	1.04	1.08	1.09	1.11	1.08	1.12	1.18	1.08	1.13	1.10	1.06	0.93
	b.	1.13	1.12	1.21	1.24	1.28	1.25	1.33	1.40	1.42	1.42	1.48	1.55	1.46
.75	a.	1.08	1.07	1.12	1.14	1.13	1.15	1.12	1.04	1.25	0.92	1.11	1.14	0.79
	b.	0.62	1.16	1.24	1.42	1.50	1.54	1.74	1.65	2.11	1.64	2.01	1.81	2.15
.90	a.	1.10	1.24	1.39	1.23	1.42	1.69	1.83	1.84	1.59	1.22	0.94	0.67	0.51
	b.	0.22	0.53	0.76	0.65	1.08	1.53	1.78	1.68	1.44	1.17	0.92	0.66	0.50
.95	a.	1.13	1.24	1.21	1.14	0.96	0.83	0.65	0.49	0.42	0.39	0.36	0.28	0.24
	b.	0.19	0.21	0.71	1.04	0.98	0.91	0.72	0.56	0.47	0.42	0.39	0.30	0.25

a = (MSE $\hat{F}_n^{-1}$)/(MSE Q_n), b = (MSE $\hat{F}_n^{-1}$)/(MSE Q_n^*)

TABLE 2. Ratios of MSEs with Uniform Kernel

(a) 50% Censoring (n=100)

$p \setminus h_n$		.03	.05	.07	.09	.10	.13	.15	.20	.25	.30	.35	.40
.10	a.	1.12	1.13	1.25	1.26	1.29	1.30	1.40	1.52	1.53	1.50	1.19	0.99
	b.	0.83	1.10	1.35	1.41	1.42	1.50	1.64	1.81	1.88	1.90	1.57	1.35
.25	a.	1.02	1.05	1.06	1.08	1.10	1.15	1.13	1.17	1.24	1.24	1.29	1.15
	b.	0.66	0.83	1.03	1.08	1.12	1.17	1.23	1.30	1.43	1.42	1.47	1.40
.50	a.	1.03	1.06	1.09	1.10	1.06	1.13	1.13	1.12	1.17	1.16	1.18	1.03
	b.	0.26	0.54	0.83	0.95	0.96	1.07	1.17	1.31	1.32	1.40	1.60	1.50
.75	a.	1.06	1.14	1.19	1.35	1.13	1.25	1.14	1.32	1.52	1.48	1.31	1.16
	b.	0.13	0.27	0.54	0.74	0.73	1.02	0.84	1.15	1.39	1.28	1.21	0.88
.90	a.	1.08	1.08	1.17	1.21	1.26	1.35	1.47	1.46	1.22	0.93	0.71	0.50
	b.	0.06	0.10	0.17	0.20	0.22	0.23	0.23	0.20	1.32	1.12	0.83	0.56
.95	a.	1.02	1.01	1.02	1.01	1.01	0.68	0.53	0.36	0.29	0.26	0.24	0.22
	b.	0.08	0.10	0.09	0.10	0.10	0.53	0.73	0.53	0.38	0.31	0.27	0.24

(b) 30% Censoring (n=100)

.10	a.	1.08	1.07	1.18	1.19	1.22	1.24	1.34	1.46	1.46	1.40	1.14	0.92
	b.	1.14	1.28	1.39	1.41	1.17	1.40	1.56	1.73	1.73	1.72	1.43	1.18
.25	a.	1.03	1.07	1.07	1.09	1.09	1.12	1.15	1.17	1.26	1.22	1.26	1.19
	b.	0.61	0.90	1.03	1.07	1.09	1.16	1.19	1.26	1.35	1.34	1.41	1.35
.50	a.	1.04	1.04	1.05	1.07	1.08	1.08	1.10	1.13	1.14	1.13	1.13	1.08
	b.	0.47	0.74	0.90	0.94	1.03	1.09	1.09	1.17	1.21	1.26	1.32	1.22
.75	a.	1.01	1.02	1.08	1.08	1.04	1.11	1.18	1.10	1.03	0.91	0.80	0.71
	b.	0.24	0.56	0.76	0.93	0.87	1.11	1.26	1.20	1.29	1.27	1.21	1.28
.90	a.	1.07	1.23	1.28	1.19	1.21	1.35	1.24	1.34	1.77	1.07	0.78	0.43
	b.	0.23	0.44	0.69	0.80	0.77	0.99	0.94	0.75	1.82	1.21	0.85	0.46
.95	a.	1.16	1.27	1.35	1.53	1.51	1.55	0.96	0.59	0.46	0.32	0.28	0.24
	b.	0.18	0.31	0.37	0.38	0.29	1.63	1.19	0.72	0.51	0.35	0.30	0.25

a = (MSE $\hat{F}_n^{-1}$)/(MSE Q_n), b = (MSE $\hat{F}_n^{-1}$)/(MSE Q_n^*)

TABLE 3. Failure Times (in Millions of Operations) of Switches

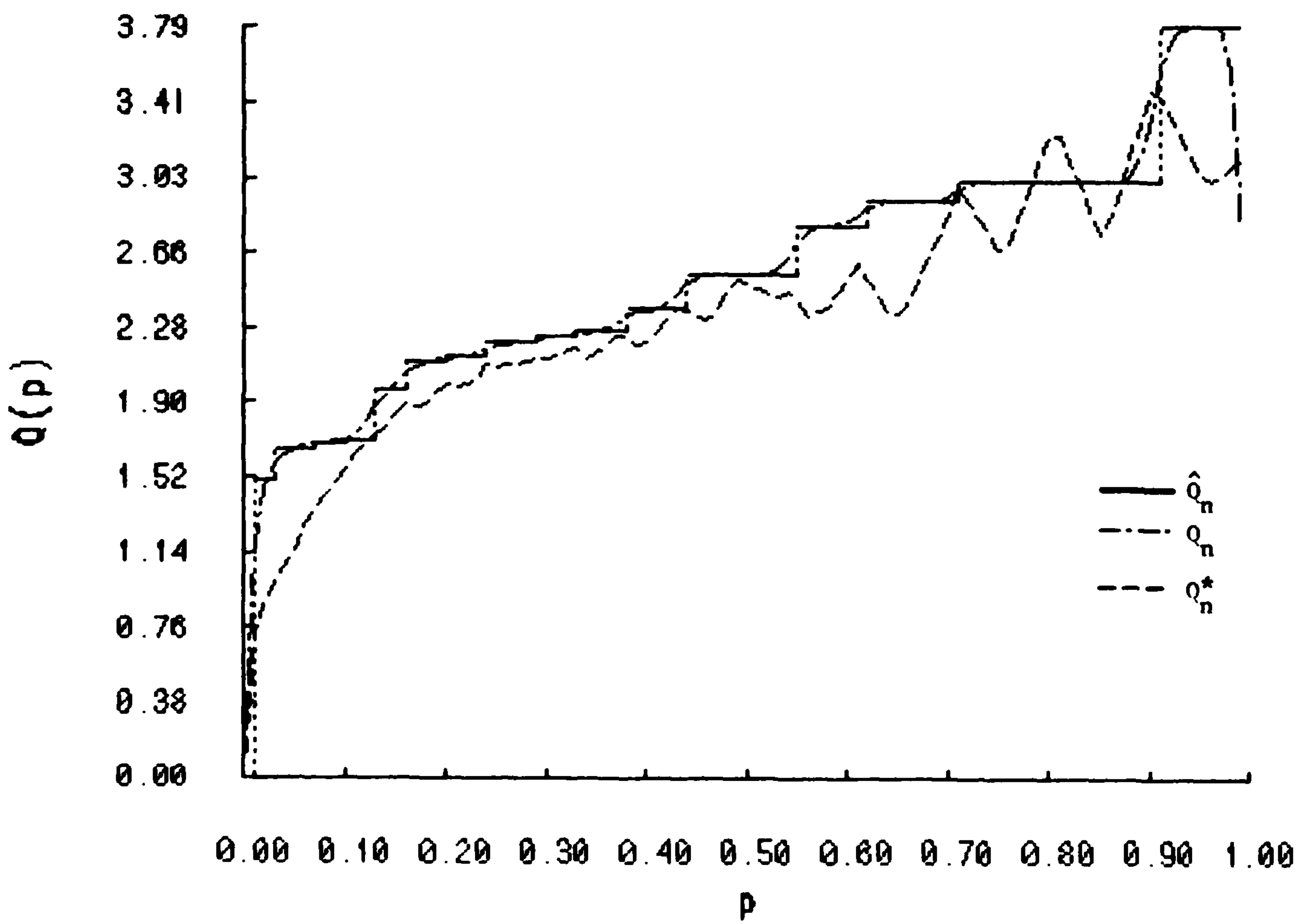
z_i	δ_i'	z_i	δ_i'	z_i	δ_i'	z_i	δ_i'
1.151	0	1.667	1	2.119	0	2.547	1
1.170	0	1.695	1	2.135	1	2.548	1
1.248	0	1.710	1	2.197	1	2.738	0
1.331	0	1.955	0	2.199	0	2.794	1
1.381	0	1.965	1	2.227	1	2.883	0
1.499	1	2.012	0	2.250	0	2.883	0
1.508	0	2.051	0	2.254	1	2.910	1
1.543	0	2.076	0	2.261	0	3.015	1
1.577	0	2.109	1	2.349	0	3.017	1
1.584	0	2.116	0	2.369	1	3.793	0

6. CONCLUSION

The kernel type quantile estimator given in this paper (and the approximate estimator) are smoother than the PL quantile function which has been used for estimation from right-censored data in the past. Based on the small Monte Carlo simulation study, however, the approximate estimator $Q_n^*(p)$ does not seem to perform overall as well as $Q_n(p)$ for $p > 0.5$, even though the two estimators are asymptotically equivalent almost surely under the stated conditions. Thus, $Q_n(p)$ seems to be the better small-sample estimator. The integrals involved in computing $Q_n(p)$ are easily calculated for the simple kernels used in Section 5.

Still under study is the problem of mean square convergence of the estimators $Q_n(p)$ and $Q_n^*(p)$ along with possible rates. Also, the choice of an optimal bandwidth sequence $\{h_n\}$ in the sense of minimum mean squared error or minimum bias is still being investigated. However, a practical choice of h_n can be obtained based on results of simulations such as those given in Tables 1 and 2. For a particular set of right-censored data, h_n can easily be changed computationally to give as "smooth" an estimate as desired.

Fig. 1: Quantile Estimators for Switch Data



APPENDIX

The proofs of Theorems 1 and 2 are presented here.

Proof of Theorem 1. First, write

$$h_n^{-1} \int_0^1 \hat{Q}_n(t)K((t-p)/h_n)dt - Q^o(p)$$

$$= \int_0^1 [\hat{Q}_n(t) - Q^o(t)]h_n^{-1} K((t-p)/h_n)dt$$

$$+ [h_n^{-1} \int_0^1 Q^o(t)K((t-p)/h_n)dt - Q^o(p)]$$

$$\equiv I + I_3.$$

Now, following the first part of the proof of Yang's (1984) *Theorem 1 with* the sample quantile function replaced by $\hat{Q}_n$, under the assumptions given, expression I can be integrated by parts to obtain

$$I = - \int_0^\infty \{\int_{F_0(x)}^{\hat{F}_n(x)} h_n^{-1} K((t-p)/h_n)dt\}dx$$

$$\pm \int_0^\infty [\hat{F}_n(x) - F_0(x)]h_n^{-1} K((F_0(x) - p)/h_n)dx$$

$$= - \int_0^\infty \{\int_{F_0(x)}^{\hat{F}_n(x)} h_n^{-1} K((t-p)/h_n)dt - [\hat{F}_n(x)$$

$$- F_0(x)]h_n^{-1} K((F_0(x) - p)/h_n)\}dx$$

$$- \int_0^\infty [\hat{F}_n(x) - F_0(x)]h_n^{-1} K((F_0(x) - p)/h_n)dx$$

$$\equiv - I_1 - I_2.$$

Using condition (F.4) and the law of the iterated logarithm (LIL) result for $\hat{F}_n$ given by Földes and Rejtö (1981), the same argument used to obtain Yang's

(1984) inequality (4) can be used to show that

$$|I_1| \le O(\max\{\frac{\log\log n}{n\,h_n}, \frac{(\log\log n)^{3/2}}{n^{3/2}h_n^2}\}) \text{ almost surely.} \tag{A.1}$$

Also, with probability one,

$$|I_2| \le \|\hat{F}_n - F_0\| \int_0^1 h_n^{-1} K((y-p)/h_n)/f_0(Q^o(y))dy,$$

where $\|\hat{F}_n - F_0\| \equiv \sup_x |\hat{F}_n(x) - F_0(x)|$, and thus, by conditions (h.1), (K.3), and (F.2) and again using the LIL for $\hat{F}_n$ (with condition (F.4)) it can be shown that asymptotically

$$|I_2| = O((\frac{\log\log n}{n})^{\frac{1}{2}}) \text{ with probability one.} \tag{A.2}$$

Finally, using conditions (h.1), (K.1) - (K.5), (F.1), (F.3), and (F.4) and Theorem 1A of Parzen (1962), it follows that

$$|I_3| = |\int_0^1 Q^o(t)h_n^{-1} K((t-p)/h_n)dt - Q^o(p)| = o(1). \tag{A.3}$$

Therefore, combining (A.1), (A.2), and (A.3), since $(\log\log n/n)^{3/4}h_n^{-1} \to 0$ as $n \to \infty$ implies that $(\log\log n)/(n\,h_n) \to 0$ and $(\log\log n/n)^{3/2}\,h_n^{-1} \to 0$ as $n \to \infty$, Theorem 1 is proved. ///

Proof of Theorem 2. For $0 \le p \le 1$,

$$Q_n^*(p) - Q_n(p) = h_n^{-1} \sum_{i=1}^n Z_i[s_i K((S_i-p)/h_n) - \int_{S_{i-1}}^{S_i} K((t-p)/h_n)dt].$$

When $s_i > 0$, that is, Z_i is uncensored, let S_i^* be an interior point of the interval (S_{i-1},S_i) with probability one so that

$$s_i K((S_i^* - p)/h_n) = \int_{S_{i-1}}^{S_i} K((t-p)/h_n)dt \quad \text{a.s.}$$

Then using condition (K.6),

$$|Q_n^*(p) - Q_n(p)| \le h_n^{-1} \sum_{i=1}^{n} Z_i\, s_i\, |K((S_i - p)/h_n) - K((S_i^* - p)/h_n)|$$

$$\le \Gamma\, h_n^{-2} \sum_{i=1}^{n} Z_i\, s_i |S_i - S_i^*|$$

$$\le \Gamma\, h_n^{-2} \sum_{i=1}^{n} Z_i\, s_i^2 \qquad \text{a.s.} \qquad (A.4)$$

Now, by the continuity of F_0 (F.1), using the definition of s_i and S_i, (A.4) can be written as

$$|Q_n^*(p) - Q_n(p)| \le \Gamma h_n^{-2} \sum_{i=1}^{n} Z_i |\hat{F}_n(Z_i) - \hat{F}_n(Z_i^-)| d\hat{F}_n(Z_i)$$

$$= \Gamma h_n^{-2} \int_0^\infty x |\hat{F}_n(x) - \hat{F}_n(x^-)| d\hat{F}_n(x)$$

$$\le \Gamma h_n^{-2} \int_0^\infty x [|\hat{F}_n(x) - F_0(x)|$$

$$+ |F_0(x^-) - \hat{F}_n(x^-)|] d\hat{F}_n(x)$$

$$\le 2\Gamma h_n^{-2} \|\hat{F}_n - F_0\| \int_0^\infty x\, d\hat{F}_n(x) \qquad \text{a.s.,} \qquad (A.5)$$

where $g(x^-)$ denotes the limit from the left at x of the function g and $\|\hat{F}_n - F_0\| = \sup_x |\hat{F}_n(x) - F_0(x)|$.

Now, since $\hat{F}_n(x) = 0$ for $x > Z_n$, $\int_0^\infty x\, d\hat{F}_n(x) = \hat{\Gamma}(Z_n)$, and by the assumptions of the theorem and the LIL for $\hat{F}_n$ of Földes and Rejtö (1981), from (A.5)

$$|Q_n^*(p) - Q_n(p)| = O((\log\log n/n)^{1/2}\, h_n^{-2}) \qquad \text{a.s.} \qquad (A.6)$$

Noting that the right-hand-side of (A.5) does not involve p, the conclusion of the theorem follows. ///

ACKNOWLEDGMENTS

The author is grateful to Winston Yang for a preprint of his paper and to Kai Yu for helpful discussions concerning Theorem 2. This research was supported by the United States Air Force Office of Scientific Research and Army Research Office under grant number AFOSR-84-0156.

REFERENCES

Breslow, N., and Crowley, J. (1974), "A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship," Annals of Statistics, 2, 437-453.

Cheng, K. F. (1981), "On Almost Sure Representations for Quantiles of the Product Limit Estimator with Applications," Research Report 87, SUNY Buffalo, Dept. of Statistics.

Csörgő, M. (1983), Quantile Processes with Statistical Applications, CBMS NSF Regional Conference Series in Applied Mathematics, SIAM, Philadelphia, PA.

Efron, B. (1967), "The Two-Sample Problem with Censored Data," Proceedings of the Fifth Berkeley Symposium, 4, 831-853.

Falk, M. (1984), "Relative Deficiency of Kernel Type Estimators of Quantiles," Annals of Statistics, 12, 261-268.

Földes, A., and Rejtö, L. (1981), "A LIL Type Result for the Product Limit Estimator," Zeitschrift für Wahrscheinlichkeitstheorie und Verwande Gebiete, 56, 75-86.

Földes, A., Rejtö, L., and Winter, B. B. (1980), "Strong Consistency Properties of Nonparametric Estimators for Randomly Censored Data, I: The Product-Limit Estimator," Periodica Mathematica Hungarica, 11, 233-250.

Gill, R. (1983), "Large Sample Behavior of the Product-Limit Estimator on the Whole Line," Annals of Statistics, 11, 49-58.

International Mathematical and Statistical Libraries (1982), IMSL, Inc., 7500 Bellaire Blvd., Houston, TX.

Kaigh, W. D., and Lachenbruch, P. A. (1982), "A Generalized Quantile Estimator," Communications in Statistics-Theory and Methods, 11, 2217-2238.

Kaplan, E. L., and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," Journal of the American Statistical Association, 53, 457-481.

McNichols, D. T., and Padgett, W. J. (1984), "A Modified Kernel Density Estimator for Randomly Right-Censored Data," South African Statistical Journal, 18, 13-27.

Miller, R. G., Jr. (1981), Survival Analysis, John Wiley & Sons, New York.

Nair, V. N. (1984), "Confidence Bands for Survival Functions with Censored Data: A Comparative Study," Technometrics, 26, 265-275.

Padgett, W. J., and McNichols, D. T. (1984), "Nonparametric Density Estimation from Censored Data," Communications in Statistics-Theory and Methods, 13, 1581-1611.

Parzen, E. (1962), "On Estimation of a Probability Density Function and Mode," Annals of Mathematical Statistics, 33, 1065-1076.

Parzen, E. (1979), "Nonparametric Statistical Data Modeling," Journal of the American Statistical Association, 74, 105-121.

Reid, N. (1981), "Estimating the Median Survival Time," Biometrika, 68, 601-608.

Sander, J. (1975), "The Weak Convergence of Quantiles of the Product Limit Estimator," Technical Report 5, Stanford University, Dept. of Statistics.

Susarla, V., and Van Ryzin, J. (1980), "Large Sample Theory for an Estimator of the Mean Survival Time from Censored Samples," Annals of Statistics, 8, 1002-1016.

Yang, S. S. (1984), "A Smooth Nonparametric Estimator of a Quantile Function," Technical Report, Kansas State University, Dept. of Statistics.